The Transcriptional Control of a Newly Identified Potassium Channel Gene, Kv3.2

by

Heather Zidow

Thesis

Presented to the Faculty of the College of Natural Sciences

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Dean's Scholars Bachelor of Science in Honors Biology

The University of Texas at Austin

December 2006

The Transcriptional Control of a Newly Identified Potassium Channel Gene, Kv3.2

Approved by
Supervising Committee:

_____

_____

_____

Acknowledgements

This paper is the product of the efforts of many people. I would like to thank the Dean's Scholars faculty, in particular Dr. Shelley Payne for her guidance on this work. I would also like to thank the principle investigator of my lab, Dr. Nigel Atkinson, for the opportunity to research in his lab.

I am especially indebted to my research supervisor, Dr. Moon Draper. Dr. Draper has spent countless hours mentoring me on research, science, and life. Some of my greatest hours of learning were spent next to him in the lab.

I would also like to thank my parents, who fostered my love of science and shaped me into the person I am today.

Abstract

The Transcriptional Control of a Newly Identified Potassium Channel Gene, Kv3.2

Heather Zidow, Dean's Scholars Bachelor of Science in Honors Biology
The University of Texas at Austin, 2007

Supervisor: Moon Draper

The completion of the human genome project revealed the existence of far more non-coding sequence (sequence that does not code for protein) than was originally anticipated. While these regions were once dismissed as "junk DNA," recent research suggests that a large portion of this material is dedicated to transcriptional control – when, where, and under what conditions a gene is expressed. The goals of this project are twofold:

1) To examine the transcriptional control of a novel potassium channel gene in Drosophila – Kv3.2.
2) To look for any commonality in the regulatory elements (the noncoding sequences that control gene expression) of Kv3.2 and the closest related gene Kv3.1.

The first part of this project was determining where in the fruit fly the Kv3.2 ion channel is made. This task required a technique called *in situ* hybridization. A probe was produced specific to the Kv3.2 gene product and a chemical reaction revealed where in the fly the probe had found its target. It was seen that Kv3.2 is solely expressed in the central nervous system of Drosophila embryos. This is identical to the embryonic expression of Kv3.1, suggesting that there may be common regulatory elements controlling the transcription of the two genes.

The second part of this project was examining the DNA sequences responsible for regulating the transcription of Kv3.2. There is currently no method for identifying these sequences directly. To do this, we must rely on the natural process of transcription to produce a gene product and then map this back to the genomic DNA. This technique was used to obtain the full-length transcriptional product of the Kv3.2 gene. The predicted transcription start site (TSS, the location where gene expression begins) was corrected and a new TSS was identified some 19,000 bases from the remainder of the gene. Knowing the location of the TSSs allows for more accurate targeting of the sequences that regulate transcription initiation. Sequence analysis was used to search for evolutionarily conserved sequences near the TSSs as potential candidates for regulatory elements.

In the post-genomic era, science faces new challenges. Whereas much of the previous effort was in locating genes themselves, the focus is now turning to identifying the regulatory elements between genes. Some of these elements for the related gene, Kv3.1, have already been identified. Similarly, future research will rely on the findings reported here to study the sequences regulating Kv3.2 expression.

Table of Contents

TABLE OF CONTENTS **6**

Preamble: Making sense of junk DNA 8

INTRODUCTION **9**

Why study transcriptional control? 9

Revelations of the genome project 9

Regulating gene expression- transcriptional control 10

Why use ion channels to study transcriptional control? 11

Ion channel basics 11

The Shaker family 13

Duplication of Kv3 14

What information is needed to understand the transcriptional control of an ion channel gene? 16

Techniques for studying the regulation of a gene 16

Characterizing the control region of Kv3.1 17

The paralog Kv3.2 18

What has been reported to date? 18

What is reported here. 19

RESULTS **20**

In situ data: Kv3.2 is expressed in the CNS 20

5' RACE data: Two new exons identified 21

Sequence analysis: Complex promoter region found 23

DISCUSSION **27**

Kv3.2 is a novel gene in the Shaker family 27

In order to study the regulation of a gene, its expression pattern must be known. 28

In order to study the regulation of a gene, its control region must be known. 29

Prior work: transcriptional regulation of Kv3.1 29

Transcriptional regulation of Kv3.2 31

IN SUMMARY **44**

MATERIALS AND METHODS **46**

RNA and DNA....46

RNA preparation....46

DNA Preparation....51

In Situ Hybridization....56

Preparation of the probe....56

Preparation of whole mount material....57

Hybridization....58

Bioinformatic Analysis....59

Software....59

REFERENCES **62**

Preamble: Making sense of junk DNA

"One man's junk is another man's treasure"

The completion of the Human Genome Project revealed a number of surprising findings. First of all, the portion of sequence coding for proteins was much smaller than expected, perhaps as little as three percent. Second, the portion of DNA that controls the transcription of genes– that is when, where, and under what circumstances a gene is expressed– is much larger and more complex than expected. In short, the sequencing of entire genomes has answered many questions but has led to many more mysteries in the process. This work uses two ion channel genes in Drosophila as a model to begin to decipher the parts of the genome involved in transcriptional control.

Introduction

Why study transcriptional control?

Revelations of the genome project

Before the completion of the various genome projects, it was largely assumed that the complexity of an organism was directly correlated to the size of the organism's genome; however, recent evidence appears to suggest otherwise. Current gene finder programs show that the plant *Arabidopsis thaliana* has at least twice as many genes as the fruit fly (Drosophila) (Rubin *et al*., 2000). Perhaps even more surprising is that there is a nearly one to one correspondence between the genes of man and mouse; fewer than 20 genes have been found to be unique to either genome (Wiehe *et al*., 2000). If the number of genes and overall size of genomes are nearly identical, wherein lie the differences between man and mouse?

The answer to this question comes from a second revelation of the genome projects—the existence of far more non-coding sequence than was originally anticipated. It was predicted, based on the average size of known genes and the overall size of the human genome, that the human genome would contain approximately 100,000 genes. Instead, it has been shown to contain little more than 30,000 (Venter *et al*., 2001). Genes, the portion of the genome that codes for proteins, comprise less than 5% of both the human and mouse genomes (Adams *et al*., 2000; Venter *et al*., 2001; Gregory *et al*., 2002). It stands to reason that the genetic differences between organisms with nearly identical coding regions, but that are so obviously different as humans and mice, would lie in this unexplored 95% of the genome.

What is the function of the remainder of the genome? It is thought that a large portion of every genome (approximately 40%) is dedicated to structural functions. However, this still leaves more than half of every genome considered intergenic sequences (sequences between the genes), that until recently were dismissed as "junk DNA" (Biessmann *et al*., 2005). Recent findings suggest that a major portion of this material is dedicated to transcriptional control—deciding when, where, and under what conditions a gene is expressed (Kraft and Horvath, 2003). These regulatory sequences contain the information that dictates how the nearly identical coding regions of man and mouse can lead to such drastically different outcomes.

Regulating gene expression- transcriptional control

Gene expression refers to the process of decoding the information contained in a gene into a particular protein. Decoding begins with transcription: the enzyme RNA polymerase synthesizes an mRNA transcript from a DNA template. In eukaryotes, this mRNA is then edited and transported from the nucleus of the cell to the cytoplasm, where it is used as a template to make a protein in a process called translation. Portions of the 5' and 3' ends of the mRNA are not made into protein, but instead determine the rate and duration of translation. These regions are called the 5' and 3' UTRs (untranslated regions).

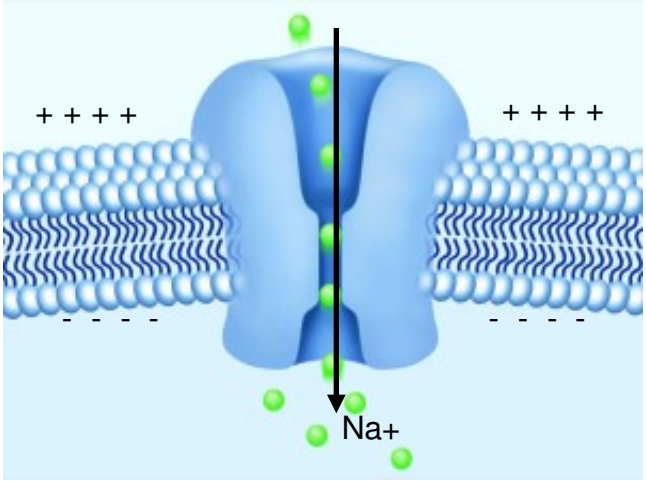
While each step of gene expression is regulated, the first step (transcription initiation) is the most important for determining which genes are expressed and how much mRNA (and consequently protein) is produced. Transcription initiation is a coordinated interaction of signal sequences present in the DNA template and the proteins that bind these sequences. One signal sequence, the promoter, helps RNA polymerase

recognize the transcription start site (TSS) of the gene. Enhancer sequences also help determine the activity of RNA polymerase at a given promoter through the binding of specific transcription factors. Some transcription factors help stabilize the transcription initiation complex, while others bind further upstream and act to stimulate or repress transcription. Thus, the level of transcription of a given gene is determined by the strength of the gene's promoter, the presence or absence of enhancer sequences, and the interaction of activator and inhibitor proteins (King, 2006).

Why use ion channels to study transcriptional control?

Ion channel basics

Ion channels are membrane proteins that allow the rapid, passive passage of ions, such as K+, Na+, and Ca++, across an otherwise largely impermeable cell membrane. These channels open due to the binding of a chemical message (ligand-gated channels) or a change in voltage across the cell membrane (voltage-gated channels). They are responsible for setting a cell's membrane potential - the voltage difference across the membrane produced by a difference in the number of negative and positive ions inside and outside the cell. Ion channels also determine a cell's excitability threshold and modulate the secretion of hormones and neurotransmitters. The combination of ion channels produced by a cell determines its unique voltage properties (Hille, 1992).



Model of a generic ion channel

Ion channels play a number of important roles in the body. They are best known for their role in the transmission of nerve impulses through action potentials. Action potentials are traveling waves of electrical excitation produced by rapid changes in the electrical potential across the cell membrane. The action potential is triggered when a stimulus depolarizes the membrane beyond the cell's excitability threshold, causing voltage-gated Na+ channels to open. As Na+ enters the cell, the membrane becomes further depolarized, causing more Na+ channels to open; thus, the process is self-propagating. Each Na+ channel is open for less than a millisecond before taking on an inactive conformation that prevents the passage of additional ions. The transient influx of Na+ is rapidly overwhelmed by the opening of voltage-gated K+ channels, which have a delayed response to the initial stimulus. K+ channels remain open until the membrane is repolarized (Alberts, 2002).

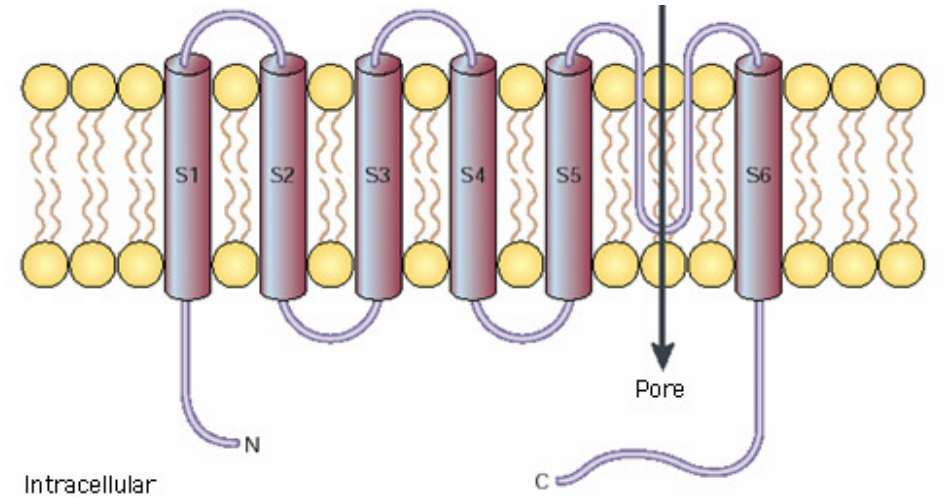
When the conduction of action potentials breaks down, there are serious biological consequences. The disease Multiple Sclerosis occurs when the body's immune system attacks the myelin surrounding nerve cells. Myelin is an electrically insulating

layer of fat, which covers the axons of most nerve cells. Without it, the speed of action potentials is greatly reduced, and affected individuals have a variety of symptoms depending on which signals are impaired.

The Shaker family

The first K+ channel gene was characterized from the Shaker mutant in Drosophila. These flies had a tremor caused by a nonfunctional K+ channel, now known as Shaker (Kamb *et al*., 1987). By searching the fly genome for sequences similar to the pore region of the Shaker protein, 3 Shaker homologs (genes of similar evolutionary origin) have been identified—Shal, Shab, and Shaw. These four genes form the Shaker family of voltage-gated potassium channels (Butler *et al*., 1989). Recently, a less confusing nomenclature has been adopted: Kv1 (Shaker), Kv2 (Shab), Kv3 (Shaw), and Kv4 (Shal).

All voltage-gated K+ channels share a similar subunit structure –six transmembrane helices (S1-S6), a pore region, and two cytoplasmic termini. The first four transmembrane helices (S1-S4) form the voltage-sensing domain, while the last two (S5-S6) form the pore domain. Between S5 and S6, there is an extracellular loop that is responsible for potassium selectivity (Jiang *et al*., 2001). Four subunits combine to form a functional channel. The subunits of one Shaker gene will not combine with the subunits of another gene (Xu *et al*., 1995); however, splice variants of a gene can form heteromultimeric channels, resulting in increased K+ channel diversity (Mottes and Iverson, 1995).



Basic potassium channel alpha subunit structure

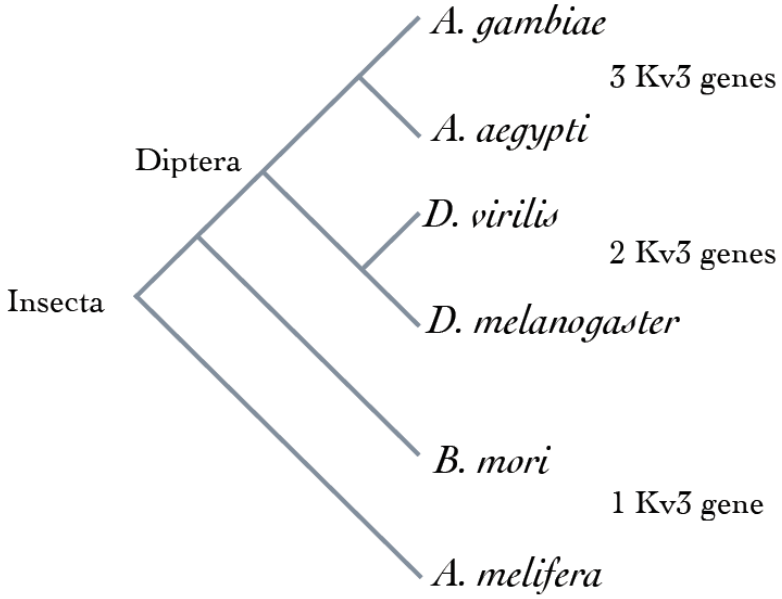
Virtually all potassium channel subunits that have six transmembrane domains conform to this basic structure. The voltage sensor is found in the S4 domain and a cannonical pore loop is found between segments 5 and 6. The Shaker family channels are all homomeric tetramers.

While the Shaker family of ion channels is derived from a common ancestor, the channels have diverged in function over time. Kv1 and Kv4 produce a fast transient K+ current, while Kv2 and Kv3 are delayed rectifiers—they remain open as long as the membrane potential is greater than the cell's excitability threshold. While all of these channels are found in neural tissue, there is preliminary evidence of differences in sub-cellular localization and cell-type specificity between them (Baro *et al*., 1994).

Duplication of Kv3

Recent research has revealed the existence of a Kv3 paralog (duplicate) in Drosophila—Kv3.2. The existence of two copies of Kv3 in Drosophila is not surprising – the moth and bee have only one copy of this gene while humans have 12 (Butler *et al*., 1989; Wei *et al*., 1990). Thus, it appears that the number of copies of Kv3 in an

organism tends to increase with the complexity of the organism. This phenomena is unique to Kv3—Kv1, Kv2, and Kv4 have not undergone such rapid duplication. This leads to the question of why Kv3 has been repeatedly duplicated.



Family tree of 6 species of Insecta and the relative number of Kv3 genes

Species from four genera show gene duplication events particular to the Kv3 locus. Both the silkworm moth (*B. mori*) and the honeybee (*A. molifera*) have 1 Kv3 gene each. Currently, only two Kv3 paralogs are found in fruit flies, while mosquitos (*Anopheles*) have 3.

Understanding the function of Kv3 is essential to understanding why it would benefit an organism to possess multiple copies of this gene. Studies expressing Kv3.1 in *Xenopus* oocytes have shown that the channel has a large conductance, activates slowly, and is non-inactivating (remains open as long as the membrane is depolarized) (Wei *et al*., 1990). This current-voltage relationship– the fact that the channel is open at the resting potential of most cells in Drosophila– suggests that Kv3 may contribute to setting the resting potential (Wei *et al*., 1990) or terminating excitation events (Rhettig *et al*., 1992).

Given the frequency of genomic mutation, it is reasonable to predict that multiple copies of Kv3 would not be conserved unless each had a distinct function. Since the Kv3 paralogs share a similar coding sequence (86% homology in the common potassium channel domains and near perfect conservation in the pore region), it is possible that this subfunctionalization is the result of differences in the regulation of their expression. The paralogs could be expressed in different cells, have different functions in the same cells, or some combination of both. The goal of this study is to gain a better understanding of the differences in the regulation of Kv3.1 and Kv3.2 in hopes of elucidating differences in their functions.

What information is needed to understand the transcriptional control of an ion channel gene?

Techniques for studying the regulation of a gene

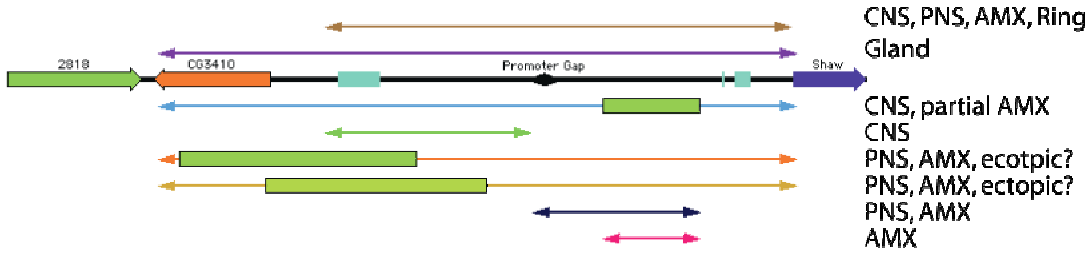
Understanding the transcriptional control of a gene begins with identifying the transcription start site (TSS) and surrounding regulatory elements of the gene. This is a search for conserved regions of DNA outside the coding sequence. Contrary to what might be expected, these sequences actually show a relatively high degree of conservation. In several species of Drosophila, it has been shown that the rate of mutation in well-described transcription factor binding sites suggests that these sequences are under functional constraint (Dermitzakis *et al*., 2003a). In addition, recent work involving whole genome comparisons between multiple mammalian sequences has shown that many previously described regulatory sequences are actually more conserved than coding sequences (Dermitzakis *et al*., 2003b). Thus, the challenge in identifying regulatory elements lies not in their lack of conservation but instead in their extreme

diversity. The best-studied promoters contain a TATA box (the sequence TATAAA) within a few hundred base pairs of the TSS. This sequence is bound by a TATA binding protein (TBP), which assists in the formation of the transcription initiation complex. However, not all promoters contain this easily identified sequence near the TSS. In Drosophila, four TBP-related proteins have been identified that help initiate the transcription of genes without a TATA box (Levine and Tijan, 2003). For genes like Kv3.2 that lack a well-described promoter sequence, computer searches are unable to locate their transcription start sites. In these situations, other techniques become necessary.

Characterizing the control region of Kv3.1

The characterization of Kv3.1 required a variety of methodological approaches. *In vitro* work was used to determine the full length of the gene. *In voltro* (*in silico*) work was used to predict the location of promoter elements. *In vivo* work was used to functionally analyze the predicted promoter elements in transgenic flies (Draper, 2005).

The regulation of Kv3.1 in Drosophila has now been described in depth. The gene has two transcription start sites, and each TSS has its own promoter elements. The region from the 5' UTR to the promoter gap is responsible for turning on the gene in the central nervous system, while the region from the promoter gap to the first coding exon controls expression in the peripheral nervous system (Draper, 2005).



CNS- central nervous system; PNS- peripheral nervous system; AMX- antennomaxillary complex

Map of the regulatory region of the shaw (Kv3.1) gene

This is an overview of the deletion construct work in the promoter region of Kv3.1. The column to the right lists the tissue specific expression for the unique reporter constructs shown to the left. The DNA sequence is denoted by the heavy black line. On this line, three genes are shown: two upstream genes (green and orange filled arrows) and the generalized exons of Kv3.1. For Kv3.1 the light blue boxes are non-coding exons and the dark blue filled arrow is the first coding exon. The colored arrows above and below the genomic are the reporter constructs with light green boxes mapping deletions.

The paralog Kv3.2

What has been reported to date?

Studies of Kv3.2 have been more limited. It was first identified in 2003 using a BLAST homology search with the Kv3.1 cDNA as the query. Several candidates were identified, and all but one were eliminated as incomplete coding regions. The final candidate was mapped at band 30B1 on the left arm of the second chromosome. This gene is designated as CG4450 (flybase ID: Fban0004450).

A previous study has reported an alignment between the two Kv3 sequences that excludes a large section of coding sequence of the Kv3.2 gene identified in this work. Based on this incomplete data, they predict that the two transcripts differ in the coding for only 12 residues. The authors perform an *in situ* assay showing that Kv3.1 and Kv3.2 have distinct expression patterns. The expression of Kv3.2 is described in early larval stages as being isolated to a small number of neural cells, while Kv3.1 is found

throughout the nervous system. From here, they extrapolate that expression must also be spatially different during the embryonic stages. They further argue that the high degree of identity of the two genes, especially in the core regions known to be important in conferring channel conductance, suggests that the channels have similar electrophysiological properties (Hodge *et al*., 2005). However, I show that the channels are expressed embryonically in similar tissues and have distinct sequences and loop structures, suggesting that they likely possess different electrophysiological properties.

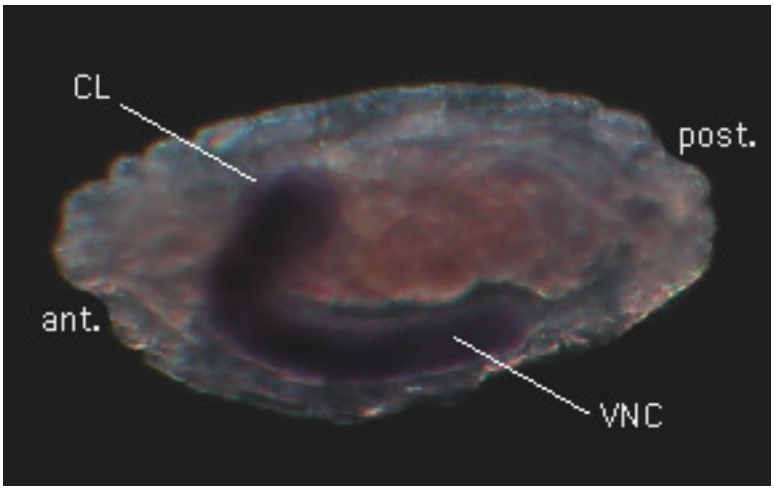
What is reported here.

I sought to answer several basic questions about this newly discovered gene: where the gene is expressed, what is the sequence of the full-length transcript(s), what is the structure of the gene in terms of exons, introns, and promoters, and how do all of these pieces of information differ from what is known about its paralog, Kv3.1?

Results

In situ data*: Kv3.2 is expressed in the CNS*

Studying the transcriptional control of Kv3.2 requires knowing which cells express the Kv3.2 gene. *In situ* hybridization was used to identify the embryonic expression pattern of Kv3.2. Expression was limited to the central nervous system.



An in situ hybridization with a probe to the Kv3.2 transcript

This is a stage 13 embryo *in situ* staining with BCIP/NBT that reveals probe hybridization (dark purple regions) to Kv3.2 transcripts. The image is a visible light micrograph. Staining is restricted to the embryonic CNS, including cephalic lobes (CL) and the ventral nerve cord (VNC).

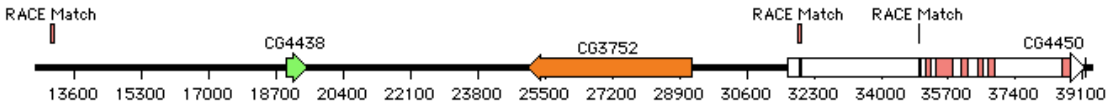
Key findings:

- Embryonic expression (stages 12-16) is limited to the central nervous system. This is different from most ion channels, but identical to the embryonic expression of Kv3.1.
- Larval expression is greatly reduced and exclusive to the CNS (data not shown). This is in contrast the Kv3.1 expression, which remains strong in the CNS and PNS (described in Discussion).

5' RACE data: Two new exons identified

Finding the genomic regions involved in regulating Kv3.2 requires knowing the full length of the gene, including its transcription start sites (TSS). These sites serve as beacons for the probable location of the regulatory sequences of the gene.

Using 5' RLM-RACE, the complete transcripts for Kv3.2 were cloned, sequenced, and mapped back to the original genomic sequences. The two RACE products yielded two new TSSs. The longest RACE product spanned three exons and included a new 5' terminal exon that was not predicted or reported from any EST database. Additionally, the TSS and first exon that was predicted has been excluded from all RACE products. Exon 2 has been altered to describe a longer exon and an extended coding region.



5' RACE results for Kv3.2

The figure maps the genomic region around Kv3.2. The numbers correspond to the number of base pairs from an arbitrary mapping point in the genome database. They are used to infer relative distances and do not represent final map locations since some gaps still exist in the genomic sequence. The predicted gene is shown as the white filled arrow with coding exons in pink. The boxes annotated as the RACE match are the sequences from the RLM-RACE mapped back to the Drosophila genome. The downstream (furthest to the right) exon is a perfect match. The green and orange filled arrows represent matches to ESTs (Expressed sequence tags). These are fragments of mRNA that have been collected and sequenced but have not been confirmed to be coding sequences. The gene represented by the orange arrow (CG3752), however, has been matched to a known protein product and a reasonable coding region has been found in the sequence.

5' RACE Continued

Key findings:

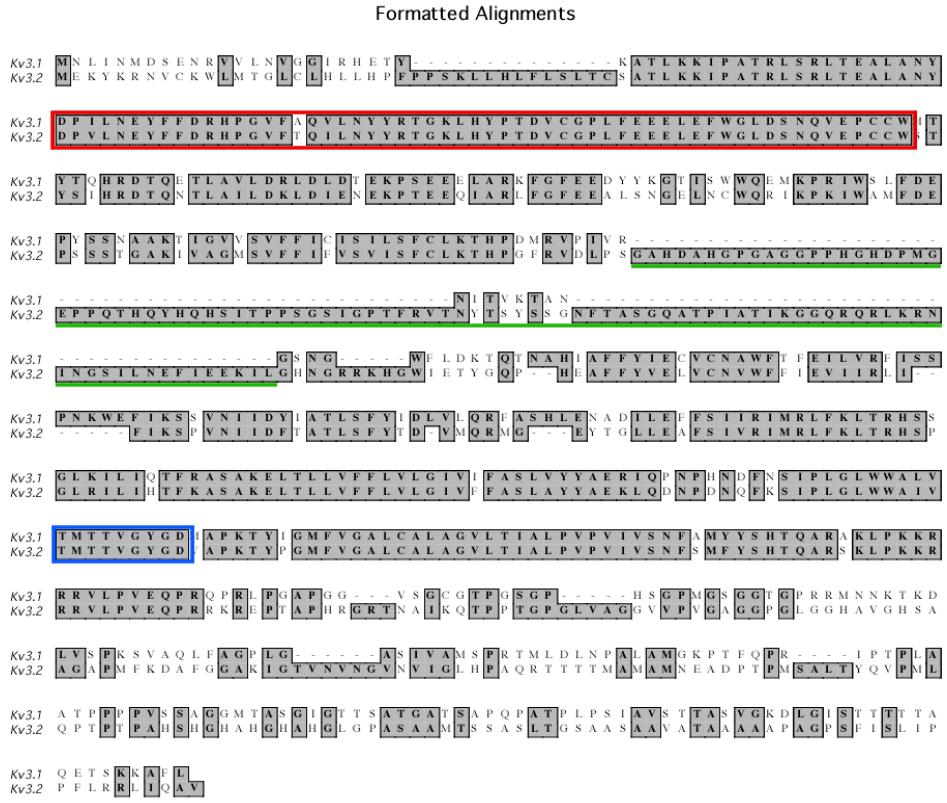
- A transcription start site was found 19 Kb upstream of the reported sequence. The intronic region contains a probable protein-coding gene and the mapped site of an EST.
- The second transcription start site is remapped to change the protein coding sequence. This TSS partially coincides with reported exon predictions. The new exon boundaries extend the coding region to include a canonical T1 domain that is common to most tetrameric potassium channels. The new sequence also produces a more likely conserved sequence for the N-terminal domain of the protein.
- A previously reported exon has been shown not to exist in the transcript.
- These new sequences correlate better with other known conserved Shaker domains than the originally predicted exons.

Sequence analysis: Complex promoter region found

While 5' RACE successfully determined approximate exon boundaries and TSS locations, we relied on sequence analysis of the Kv3.1 and Kv3.2 proteins and of the noncoding sequences surrounding Kv3.2's TSSs for finer scale analysis.

The protein

A comparison of the peptide sequences of all the Shaker cognates reveals that Kv3.1 and Kv3.2 are the most related genes. They share a greater than 50% sequence conservation, with some domains entirely preserved.



Clustal-W Alignment of Kv3 Polypeptides.

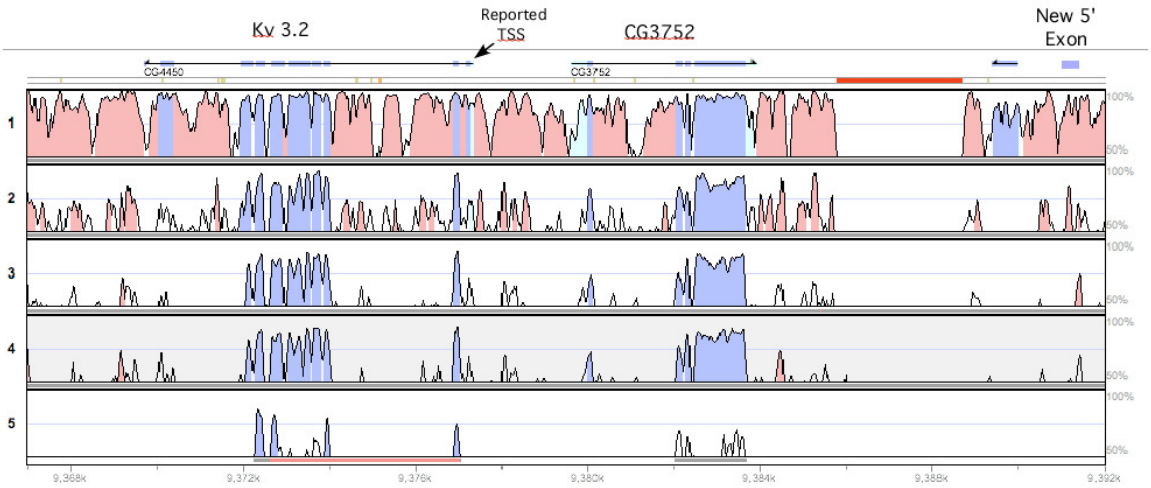
This is a slow Clustal-W alignment. Gap penalty was maximized. Grey filled boxes denote conservation of sequence (identity) and amino acid class (R group function). Colored annotations are described below. Alignment scores in the final third of the peptide are difficult due to splice variants.

An alignment of the Kv3.1 and Kv3.2 polypeptides reveals a strong degree of conservation. The pore region sequences (blue box) are identical and the tetramerization (T1) domains (red box) are nearly identical as well. There is also high conservation in each transmembrane domain. The major difference between the proteins lies in Kv3.2's long extra-cellular loop between domains I and II (green line). This domain has been shown to be the portion of the subunit responsible for alcohol sensitivity in mice and the Kv3.1 ortholog, which is lacking this domain, does not respond to alcohol treatment (Shahidullah *et al*., 2003). This pore loop domain was confirmed using RT-PCR techniques and an *in situ* probe specific to the loop and the second transmembrane segment. Previously reported sequences do not show this 90 amino acid loop in Drosophila.

The regulatory region

While the coding regions of Kv3.1 and Kv3.2 are quite similar, the non-coding regions show no conservation of sequence. Even the sequences of the first exons– those at the transcription start sites (TSS)– are not preserved. Without a TSS to serve as an anchor point to look for similarities in the regulatory sequences, there is little hope of finding conserved transcription factor binding sites.

The alternative is to compare the sequences between closely related species of the same genus. The split between the two Kv3 genes occurred before the genus Drosophila evolved. Therefore, some regions of the control sequences may have been conserved between members of the genus.



Multispecies alignment (SLAGAN) in the genomic region around Kv3.2.

The sequence from *Drosophila melanogaster* (reference sequence, map features at the top) is compared to 5 species. The numbered (left) graphs represent relative degree of sequence similarity between the reference sequence and the assembled contigs from other genomes.

The numbered graphs 1-5 represent the reference sequence aligned against the following species in order: *D. yakuba, D. pseudo obscura, D. virilis, D. mojavensis*, and *A. molifera*. Details of the figure are described below.

An alignment of four species of Drosopholids and the honeybee to the genomic sequence of *Drosophila melanogaster* reveals a high degree of conservation in the coding region (purple peaks) of Kv3.2. Pink regions denote areas of conservation greater than 70% over a 50 base pair window. The red bar signifies a large gap in the known sequence alignments. Included in the plot is the gene CG3752, showing that it is extant through all species including *Apis melifera*. In the right hand portion of the plot is a computer-predicted ORF fragment found only in melanogaster and its close cousin yakuba. The fact that this sequence is not conserved despite being labeled as coding indicates that it is likely not an expressed gene. Of particular interest, the reported TSS is less conserved than either of the new TSSs described here, lending credence to the argument that the reported TSS is invalid. Lastly, some narrow regions of intronic sequences are more conserved than the non-coding exons. These are candidates for

transcription factor binding sites, which would be conserved due to their functional significance.

Key Findings:

- The coding regions of Kv3.1 and Kv3.2 are highly conserved.
- Kv3.2 possesses a complete T1 domain with the extension of exon 2.
- The extra-cellular loop found in mammalian orthologs is also present in the Drosophila transcript.
- All other critical channel domains were mapped.
- The map of the regulatory region was expanded and another gene was found to exist within this control region.
- The RACE data extending exon 2 is conserved in other species of fruit flies, while the predicted first exon is absent from other species.
- The new 5' exon identified with RACE is conserved at its 5' end.
- Avid/VISTA plots reveal likely candidates for transcription factor binding sites.

Discussion

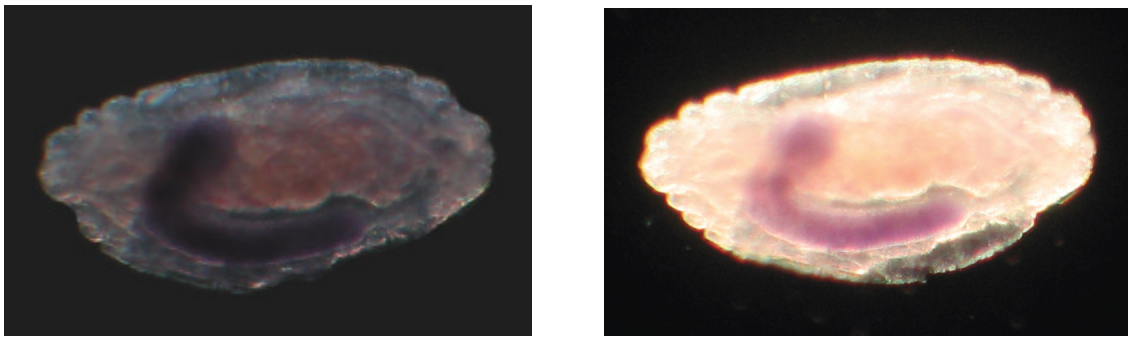
Kv3.2 is a novel gene in the Shaker family

Upon the release of the completed Drosophila genome, an opportunity was afforded to search for homologs to known genes within fruit flies and other species. While a search with Shaker and the homologs Kv2 and Kv4 revealed no other members of the Shaker cognate family, a new gene designated CG4450 was found to be the closest relative to Kv3. The Shaker cognates Kv2, 3, and 4 had been identified using molecular biology techniques in a search for genes similar to Shaker (Kv1). A fragment of the Shaker transcript was used to probe a cDNA library of adult mRNAs and these three cognates were found due to high conservation of sequence surrounding the pore region. One of the limitations of probing cDNA libraries is that transcripts of low abundance are likely to be missed. Since Kv3.2 appears to be expressed in a very small region of the CNS in adult flies (Hodge *et al*., 2005), only with the release of the whole genome sequence was this gene finally identified.

Once identified, this gave rise to a series of questions that fueled this research. These included: What is the function of this novel potassium channel? Where is it expressed in the fly? Since it has a strong sequence similarity to Kv3.1, could it be used to compare and contrast the sequences that regulate the expression of these two genes? And lastly, is this gene similarly duplicated in other species?

In order to study the regulation of a gene, its expression pattern must be known.

An *in situ* assay is a molecular biology technique to identify the localization in cells or tissues of gene products. Since the Kv3.2 protein has not been isolated, no antibodies exist to perform a histochemical assay. For this reason, an *in situ* was conducted using a cDNA probe. A 700 base pair probe was generated to regions unique to the Kv3.2 transcript. While all the Shaker cognates have similar pore regions, the first and second transmembrane segments are not highly conserved.



In situ hybridizations using probes specific to Kv3.2 (left) and Kv3.1 (right)

Comparison of Kv3.2 (left) and Kv3.1 (right) stage 13 embryos. Expression is largely limited to the CNS, which includes the ventral nerve cord (VNC, long structure at the bottom) and the cephalic lobes. By stage 17, Kv3.1 expression is found in the PNS.

Kv3.2 expression was found to be very similar to that of Kv3.1 in embryonic stages 12-16. However, while Kv3.1 expression remains strong in the CNS and PNS in larvae, expression of Kv3.2 is greatly reduced and exclusive to the CNS. In embryos, expression of Kv3.2 is particularly strong in the ventral half of the VNC. With maturation, this structure shortens to form the CNS, leaving an afferent and efferent PNS neural network. Concurrent work showed preliminary results in an adult *in situ*, wherein

the expression of Kv3.2 is reduced to a few neurons in the cephalic lobe (Hodge *et al*., 2005).

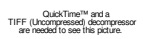
In situ hybridizations using probes identifying transcription products are not definitive evidence for the presence of actual channels. However, oocytes injected with transcripts of Kv3.2 homologs from many species have produced functional channels with measurable currents (Attali *et al*., 1993). Therefore, it is reasonable to assume that a native Kv3.2 protein is extant in Drosophila. Efforts to isolate antibodies to this protein are ongoing. Since the two Kv3 siblings are initially expressed in similar tissues, it was hypothesized that the regulation of these two genes is under the control of similar sequences. If the transcriptional control regions of the two genes could be identified, this would provide an opportunity to identify CNS-specific elements in these regions. This project originated on the heels of work that defined the control region and discrete regulatory elements in the Kv3.1 gene.

In order to study the regulation of a gene, its control region must be known.

Prior work: transcriptional regulation of Kv3.1

The regulation of Kv3.1 in Drosophila has been described in depth. In general, it is known that the lower (3') half of the regulatory region of the Kv3.1 gene controls expression in the peripheral nervous system while the upper (5') half controls expression in the central nervous system. Dissection of the Kv3.1 control region required a variety of approaches. A 5' RACE resulted in the description of two new transcription start sites and three noncoding exons. Sequence analysis showed small regions that were

particularly conserved throughout Drosophilid evolution. *In vivo* assays were used to determine the function of these sequences. This work is shown below.



Summary of Kv3.1 work

At the top of the figure is an annotated genomic sequence in the Kv3.1 control region. Below this map is a trace showing conservation of sequence between several Drosophila species. Coding regions are in purple; noncoding regions that are highly conserved are in pink. The middle drawings are from the Atlas of Drosophila Development showing the neurological features of a stage 13 Drosophila embryo. The bottom portion of the figure shows micrographs of the four transformant lines produced using the construct maps shown.

Key Findings:
A) There were two promoters (transcription start sites) mapped with 5' RACE.
B) Schematic of the CNS (purple) and PNS (blue) of the embryo.
C) Overview of Kv3.1 transformant lines.
    Aqua construct [top left] deletes PNS promoter (green box)
        Yields only CNS expression
    Lime construct [bottom left] inserts only CNS control region
        Yields expression only in CNS
    Orange construct [top right] deletes CNS control region (green box)
        Yields only PNS expression.
    Pink construct [bottom right] inserts only a portion of the PNS region
        Yields expression in a portion of the PNS, the antennomaxillary complex.

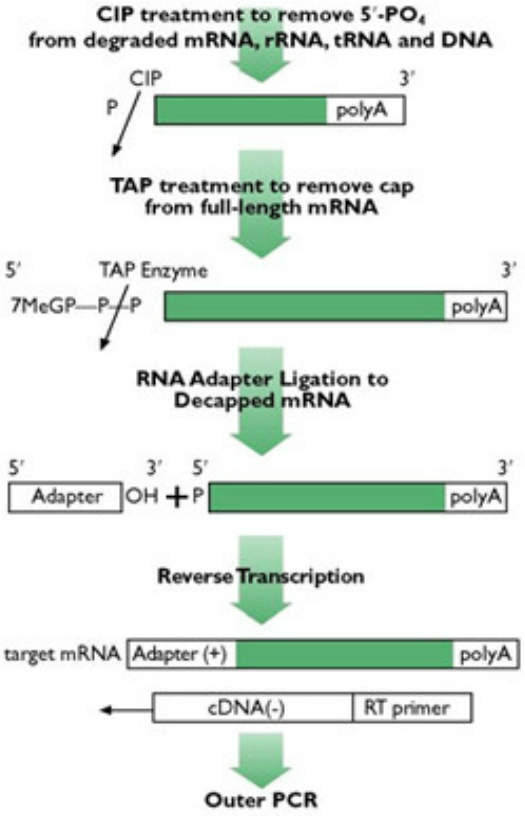
In order to begin to identify the regulatory elements controlling Kv3.2 expression, the transcriptional control region of the gene had to be identified. This could then be used to compare with the data from Kv3.1 (see above). The path to this begins with a 5' RACE (Rapid Amplification of cDNA Ends), which shows where transcription of a gene begins. Once the transcription start sites are known, the search for regulatory sequences can commence.

Transcriptional regulation of Kv3.2

5' RACE assay

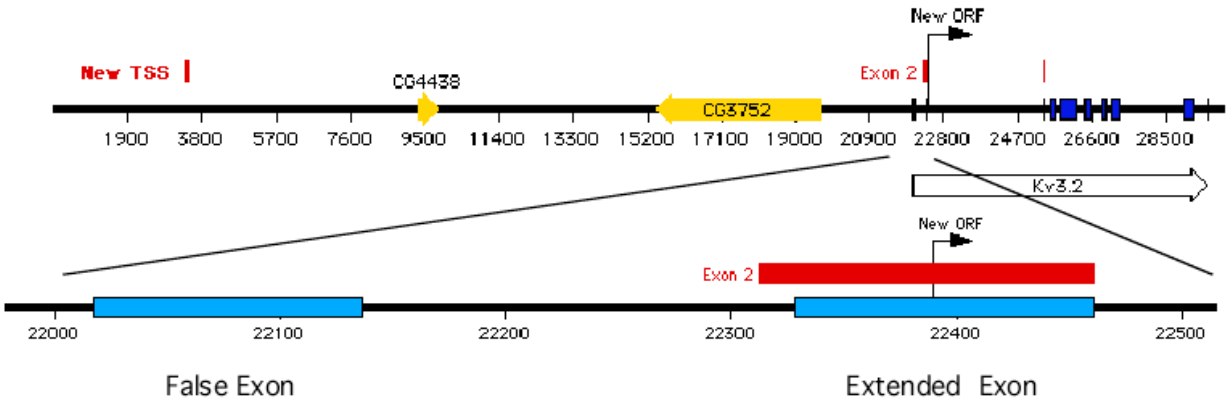
RACE is a way of finding the complete mRNA of an expressed gene. One of the consistent problems with mRNA analysis is cloning the complete message up to its 5' end. The 3' terminus is easily identified due to the location of a stop codon and the subsequent polyadenylated tail; however, the 5'–most methionine and 5' noncoding exons cannot be known without sequencing a complete message. 5' degradation of message often leads to shortened clones, and historically transcripts were only reported to a probable full-length based on coding sequence and comparison to known protein sequences. I used a version of RACE called RLM (RNA Ligase Mediated)-RACE,

which attaches an RNA fragment of known sequence to the 5' end of only complete messages. Following PCR and cloning reactions, the complete 5' end can be identified.



I used poly-A selected embryonic mRNA as the source material for my RACE. This was treated with Calf Intestinal Phosphatase (CIP) to eliminate partially transcribed and degraded messages from the pool of candidates for the subsequent adapter ligation step. At this point, only two species of RNA existed in the extract: dephosphorylated incomplete messages and 5' capped complete messages. Next, the unique Tobacco Acid Phosphatase (TAP) enzyme was used to cleave the pyrophosphate bond of the 7-methyl cap, leaving only complete messages with a 5' phosphate available for the ligation of the adapter RNA fragment. From this point forward, a normal 5' RACE using an RT reaction and PCR was conducted. A further step to increase specificity involved nested PCRs.

The 5' RACE of Kv3.2 identified a new 5' exon far upstream from the predicted transcription start site. This new TSS greatly expands the search area for regulatory elements. Additionally, an exon (false exon) that had been predicted and reported is shown not be part of the genuine transcript. Lastly, exon 2 was extended, establishing the location of the second TSS and completing a conserved 5' terminal domain known as the T1 domain, which is common to the vast majority of tetrameric potassium channels.



Map of the RACE results

Top map shows the genomic region surrounding Kv3.2 over a span of 35 kb. Dark blue boxes represent coding exons. The yellow filled arrows are the genes found within the first intron. The red boxes above the line are the mapped RACE product, which spans three exons. The new protein TSS is labeled with the black arrow above the map line.

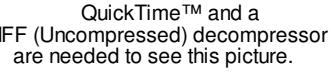
The lower map is an enlargement of the region around the first two predicted exons. The RACE result (red box) shows the 3' boundary of exon 2 is correct while the 5' boundary has been extended. The predicted exon 1 was never seen in any of the RACE products and any RACE with a primer designed to this specific exon failed.

These results possibly explain the great difficulty that other researchers have had trying to generate probes to this message from the 5' end. Several labs have produced constructs from the predicted transcript, which have failed to express a functional protein in oocytes. This new sequence may permit expression of the full-length sub-unit in oocytes for physiological studies. Previous studies have relied on chimeric channels with other Shaker-family alpha sub-unit gene products for such studies (Smith-Maxwell, 1998).

The discovery of a new TSS 19 kb upstream of the coding region is advantageous in the effort to map out the transcriptional control region, but also illustrates one of the critical issues in promoter studies– that any intervening genes will complicate the description of transcription factor binding sites. There is no way to know whether any sites identified through molecular or computational techniques regulate the gene of

interest or these neighboring genes. Classically, promoters and their enhancers were described as being proximal and upstream of transcription start sites. While this has quite often been the case, the Kv3.1 studies show that regulatory sequences can lay downstream of the TSS as well as upstream within a single control region. What is more, Ohler *et al* have found regulatory sequences as far as 100 kb away from the transcription start site in their whole genome studies in Drosophila (Ohler *et al*., 2002). This presents the problem of locating candidate sequences for molecular analysis of complex control regions.

It was my intent to compare the control regions of Kv3.1 and Kv3.2 and use what is known of the Kv3.1 promoters to target promoter elements in Kv3.2. However, there exist great differences between the two gene structures. While all members of the Shaker cognate family have a two promoter structure, the size and presence of other genes included in this intronic region makes the Kv3.2 regulatory region a particular challenge to study. Below is a comparison of the gene structures of all four Shaker cognates.



General Structure of the Promoter Regions of the Shaker Cognates

This figure shows the general layout of the noncoding 5' exons in the intergenic sequences upstream of the Shaker cognate genes. In general, each gene has two TSSs that are from 3,000 to 30,000 bases apart. Only Kv3.2 has a known protein-coding gene included in the region between the two TSSs. The maps are not to scale to one another.

Blue boxes are Shaker cognate exons. Green arrows are computer predicted genes. Gray arrows are known neighboring genes, and yellow arrows are the 'intronic genes' found in the Kv3.2 regulatory region. Red arrows denote transcription start sites.

Sequence analysis

Molecular techniques leading to transformant lines are critical to identifying functional elements in the promoter region of a gene. However, these techniques are also time consuming, and it is desirable to narrow the scope of the search region for regulatory elements through other means. Towards this end, determining the boundaries of the coding region is crucial to mapping the regulatory region. Previously, the initial methionine was inferred from homologous sequence comparisons. Now much more is known about basic ion channel structure and looking for complete critical domains can be used to determine the probable start of the coding region.

Once the extent of the coding region is known, locating the TSSs is the next vital step to setting the boundaries of the transcript. Successfully mapping a complete transcript to the genomic DNA via RACE is necessary to reveal the location of putative promoters. However, even once the TSSs have been identified, the promoters may not be readily apparent. This has been illustrated in a scan of the complete Drosophila genome showing that less than 40% of the known transcribed genes have a canonical TATA sequence. In addition, the scan revealed ten basal motifs that exist in the predicted range of the TSSs of more than 9,000 genes (Ohler *et al*., 2002).

In the course of conducting sequence analysis, both coding and noncoding sequences were examined. The former provides information about the protein itself and the evolutionary conservation of its function, whereas the latter is an avenue to understanding the regulation of the gene.

The protein

The peptide sequences of all the Shaker cognates were compared and all had the general pattern of 6 transmembrane segments and the signature GYGD pore domain found in all potassium channels. Missing from the reported Kv3.2 peptide sequence was a complete tetramerization domain. The T1 protein domain is believed to guide the tetramerization of the alpha-subunits in the complete channel. It consists of two regions, a common region that is likely used to raft the proteins on the cell surface and an "identity" domain that dictates the assembly of homomeric channels, excluding the subunits of similar genes (Tu *et al*., 1996). The extended exon, found using RACE, shifts the reading frame and allows for a full-length canonical T1 domain.

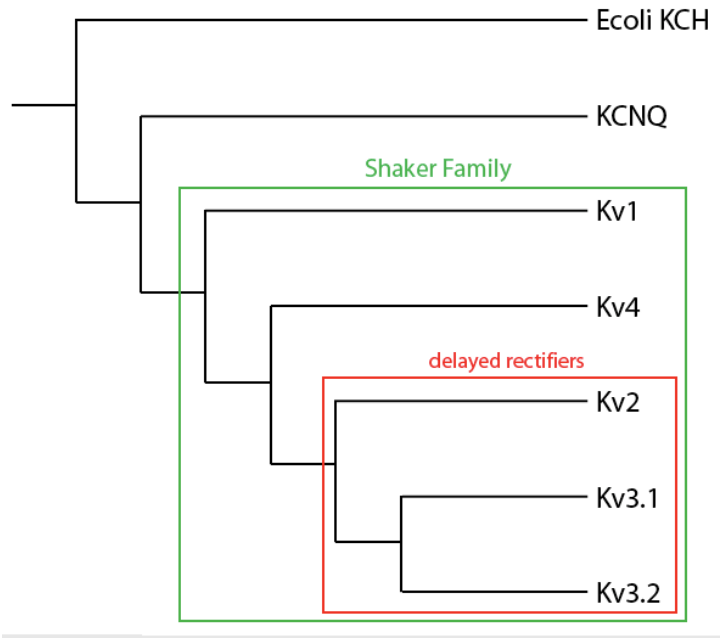
Aligned T1 tetramerization domain sequences

The bottom line (Common T1) shows the T1 domain sequence that is conserved across many species, from invertebrates to mammals. The predicted T1 sequence for Kv3.2 is not a good match to this sequence or to the domain in Kv3.1. The Kv3.2 extended exon found using 5' RACE is identical to the T1 domain in Kv3.1 and is a better match to the consensus T1 domain.

Another new sequence described in the molecular studies is a large extra-cellular loop between transmembrane segments I and II. Loops that span between transmembrane segments are highly variable in all ion channels. They are often the site of protein modification, including phosphorylation, acetylation, and glycoslyation. This is typically the primary distinction between channels of similar structure but varied function. For example, a calcium-binding domain has been identified on the intra-

cellular loop of calcium-activated potassium channels (Yu and Atkinson, 2006). Several potassium channels have been shown to be sensitive to the presence of alcohol and the binding site on at least one channel has been identified in a loop domain. The ortholog to Kv3.2 in mice in particular, has been shown to be the most alcohol sensitive of the Shaw-related mouse potassium channels (Shahidullah *et al*., 2003).

Alignment of the protein sequences of the entire Shaker cognate family confirms that Kv3.2 is more closely related to Kv3.1 than the other members. Furthermore, a generalized alignment with all of the cognates clusters Kv2 along with Kv3.1, and Kv3.2. While the function of Kv3.2 has not been ascertained, it is reasonable to assume that it will be similar to that of Kv3.1 and Kv2. Both of these channels are delayed rectifiers that are largely responsible for establishing the resting potentials of the cells in which they are expressed. Kv3.1 is found in neural tissues, as is Kv2; however, Kv2 is restricted to axon termini (Wei *et al*., 1990), while Kv3.1 is expressed throughout the cell. Expression of Kv3.2 in adult flies has only been isolated to the scale of a few whole cells in the CNS (Hodge *et al*., 2005).



Cladogram of the Shaker potassium channel family in Drosophila.

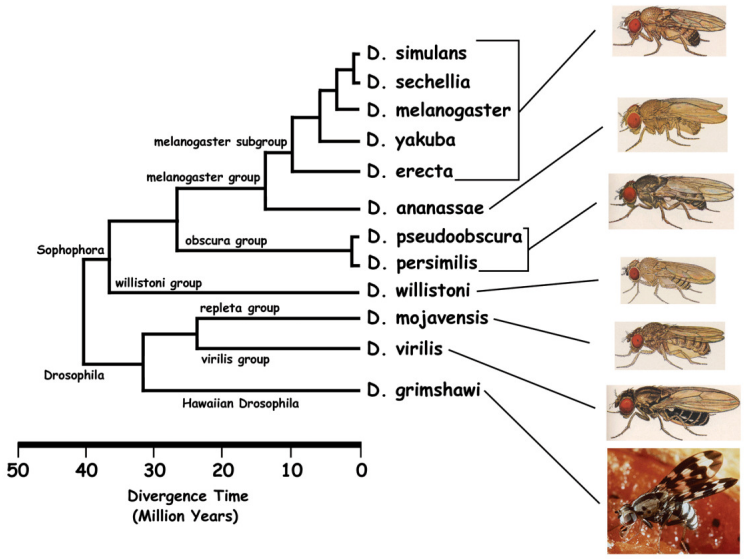
The central coding region (excluding N and C terminal domains) of 7 potassium channels were compared. The evolutionary relationship of these sequences reveals distinct subgrouping of ion channels. The Kv1-4 channels are in the Shaker cognate group. A second family of voltage-gated channels in the fly are designated KCNQ for their similarity to the human KCNQ genes.

The regulatory region

Evolutionary selective pressures on the protein coding regions have maintained conservation between the two sequences at greater than 65%. However, conservation of the control region is far less, and a direct comparison of the Kv3.1 and Kv3.2 control regions is not informative. Both Kv3.1 and Kv3.2 are found in other insect species including mosquito (*Anopheles gambiae*), silk worm moth (*Bombix mori*), and honeybee (*Apis melifera*). Therefore, the duplication event that created a second Kv3 gene must have happened prior to these speciation events, believed to have occurred more than 120 million years ago. Because the sequences in the regulatory regions of the same gene are

more likely to be conserved between species than between the two different Kv3 genes, I chose to compare the noncoding sequences surrounding Kv3.2 between closely related species of Drosophilids.

A multi-species alignment was performed using the Slagen and AVID algorithms. A total of five Drosophila genomic sequences were aligned (*D. melanogaster, D. yakuba, D. pseudo obscura, D. virilis, D. mojavensis*), with the honeybee sequence used as an outlier.

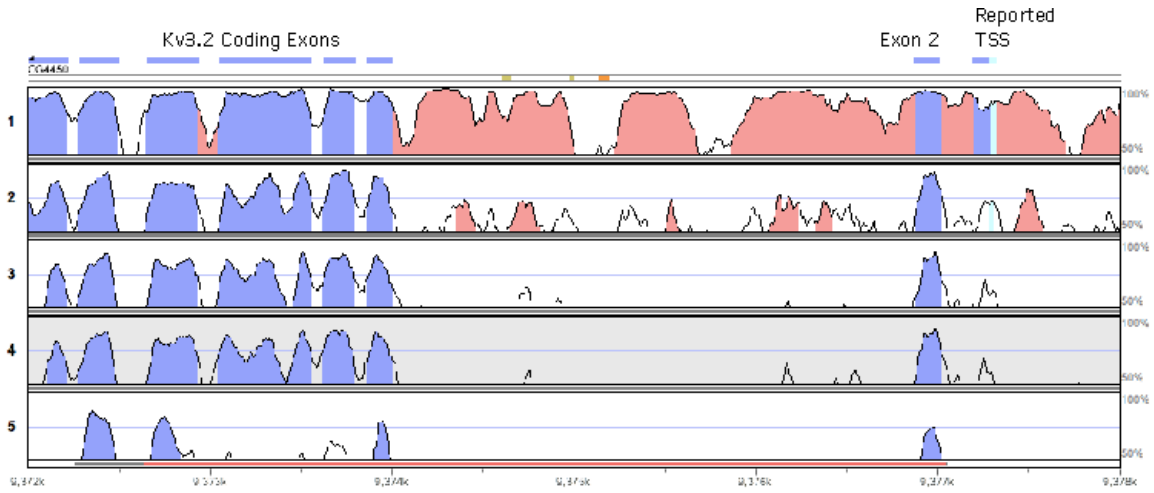


Cladogram of the Drosophila genus showing evolutionary relationship between 12 species of fruit flies.

The species shown are all candidates for genomic sequencing. With the exclusion of melanogaster and pseudo obscura, all projects are ongoing and some gaps still exist in the sequences. For this reason, only the sequences that were near completion in the region of Kv3.2 were used in the genomic alignments.

Slagen and AVID search algorithms are specifically adapted to the nature of genomic sequences. Unlike BLAST search and Clustal alignment algorithms, which are used to align coding sequences of DNA, Slagen and AVID use the BLAT search paradigm, which is capable of aligning small sequences (less than 100bp) to large fields

of genomic sequence by accommodating large gaps in sequence without excessive penalty. The second step is to use AVID or Slagen to combine the BLAT alignments and display them with the VISTA program, which plots the degree of sequence conservation between the species over a specified window size. The plots are generated by scoring the conservation over the window and plotting this value at the midpoint of the window on the graph. Regions exceeding a threshold of 70% conservation are marked in pink, and coding regions are marked in purple. As expected, the coding regions of the genes are highly conserved, whereas intronic and intergenic regions are not. However, localized peaks of conservation of short sequences are likely candidates for functionally significant elements.



VISTA plot of 5 species alignment of the coding region of Kv3.2

This AVID run used parameters that emphasized conservation of coding region and exon boundaries,

The sequence from *Drosophila melanogaster* (reference sequence, map features at the top) is compared to 5 species. The numbered (left) graphs represent relative degree of sequence similarity between the reference sequence and the assembled contigs from other genomes.

The numbered graphs 1-5 represent the reference sequence aligned against the following species in order: *D. yakuba, D. pseudo obscura, D. virilis, D. mojavensis*, and *A. molifera*. Details of the figure are described below.

This plot focuses on the conservation of sequence around the coding region of Kv3.2. Here the reported TSS is not very well conserved in comparison to the other exons. However, the sequence located 5' of exon 2 is relatively conserved, supporting the RACE findings that expand this exon an additional 80 bases. Exon 2 is even conserved in the bee genome. The gap in the coding region of the bee corresponds to the long extracellular loop that exists in the fruit fly between transmembrane segments 1 and 2. These loops are typically more variable than transmembrane segments.

Another potential approach to identifying functional elements in promoter regions requires information that is not yet available, in particular a catalog of transcription factor binding sites. The TransFac Database is relatively nascent, with fewer than 100 candidate transcription factor binding matrices out of a theorized 3000-4000. The regulatory region of Kv3.2, both proximal and the entire 19kb region, was searched using the Transfac Database. Preliminary results revealed no hits of significant p values under stringent conditions (greater than 85% match to core Transfac matrix). Conversely, low stringency searches, which allowed for variance in the core sequence match, revealed too many hits to reasonably discern a pattern. These search results suggest that the Transfac Database has insufficient information and database depth to be of any practical use at this time.

These findings illustrate that despite knowing the full genomic sequence, understanding the function of the transcriptional control region can be stymied by the complexity of the genome itself. In whole genome studies, the vast majority of the genes were found to be non-overlapping. Kv3.2, however, has a promoter region that is both particularly large and complex, which includes the coding region of another gene.

Several hurdles need to be surpassed before further work can be done. These include the generation of mutants and immuno histochemical reagents, and the completion of the sequence of the other Drosophila genomes in the region of Kv3.2.

In Summary

The findings of the sequence analysis were in concert with molecular and *in situ* data. The description of the regulatory region of Kv3.2 revealed that it is more complex than that of its paralog such that it may be an intractable problem for direct *in vivo* analysis using expression constructs until regulatory sequences can be more discretely defined. The inclusion of another coding region within the first intron of Kv3.2 confounds the association of any identified promoter elements to a particular gene. Even so, it is interesting to find that a particular member of a family of potassium channels found in most eukaryotes has undergone multiple gene duplication events. This raises many questions of subfunctionalization and variability in the regulation of related genes.

Here I have shown that the Kv3.2 paralog is expressed in a similar fashion to Kv3.1 in the embryonic stages, that the gene structure on the other hand, is significantly different, and that the promoter region bares little if any similarity to the paralog. These three findings support the general notion that while homologs may be quite similar in their sequence and possibly their function, their regulation can be quite different. It is this regulation that gives rise to the variances found between species more so than the genes themselves.

ZZZZZZZ

One of the most promising treatments for cyctic fibrosis pursued in the 1990s was gene therapy. In this approach, a functional copy of the affected gene, a chloride transporter, is carried to the effected tissues using an adenovirus vector. In patients, promising results were seen as the transgene successfully produced the needed protein and alleviated patients' respiratory symptoms. However, serious complications arose that

resulted in the death of a patient. The issue was that while the protein product was generated, there was no endogenous control over the proper amounts of protein to be expressed. As a result, over-expression of the targeted gene led to tissue necrosis and death. The solution to the problem would be to include the DNA sequences that normally control this gene's expression in the transgene. Unfortunately, we currently don't know how to identify the transcriptional control regions of a gene satisfactorily. This work represents a few tentative steps in this direction.

Materials and Methods

Many of the techniques used in this work are based upon standard protocols. However, most were modified to accommodate particular challenges.

RNA and DNA

RNA preparation

Total RNA

Modified RNA preparation techniques based on the One-Step method (Sambrook) were used. Drosophila represent a challenge in that the cuticle mass makes separation of proteins from nucleotide material critical. To this end, often two or three cycles of acid-phenol extraction and subsequent precipitations proved effective.

The general protocol.

Reagents:

Guanidine isothiocyanate (Sigma)

1M sodium citrate, pH 7 (DEPC-treated, autoclaved)

Sarcosyl (N-lauryl sarcosine, Sigma)

b-mercaptoethanol (Sigma)

2M sodium acetate, pH4 (DEPC-treated, autoclaved)

3M sodium acetate, 100mM magnesium acetate, pH 5.2

Absolute ethanol

Propan-2-ol (isopropanol)

70% ethanol (made with DEPC-treated, autoclaved water)

0.5% SDS (made with sterile, DEPC-treated water)

Denaturing Solution- 4M guanidine isothiocyanate, 25mM sodium citrate, pH7, 0.5% sarcosyl, 100mM ß-mercaptoethanol.

(Denat. Sol. can be made and stored at 4°C without ß-mercaptoethanol for several months. ß-mercaptoethanol should be added to 100mM immediately prior to use.)

1) Tissue is homogenized as rapidly as possible, at 4°C, in solution D (500ul per 50mg tissue) with an eppendorf pestle homogenizer until a smooth, lysed, homogenous suspension is obtained.

2) Add 50ul 2M sodium acetate, pH4.0 and mix vigorously.

3) Add 500ul phenol and mix vigorously.

4) Add 100ul chloroform, mix vigorously and incubate on ice for 15 minutes.

5) Centrifuge mixture at 10,000g for 10 minutes in a microfuge at 4°C.

6) Remove upper, aqueous phase to a clean, sterile, DEPC-treated eppendorf tube. After centrifugation, RNA is present in the aqueous phase while, due to protonation at the acidic pH used, genomic DNA is partitioned into the phenol phase.

7) Extract the upper aqueous layer with an equal volume phenol/chloroform and centrifuge as before. Repeat the extractions until no interface material is seen.

8) Precipitate the aqueous phase by the addition of an equal volume (500ul) of propan-2-ol. Incubate at -20°C for 20 minutes.

9) Pellet RNA by centrifugation at maximum speed in a microfuge for 10 minutes.

10) Wash the RNA once in 70% ethanol and vacuum dry.

11) Re-dissolve in 200ul 0.5% SDS at 65°C.

12) Extract with an equal volume (200ul) of phenol/chloroform as above. Repeat until no interface material is visible.

13) Precipitate pure RNA by the addition of 20ul 3M sodium acetate, 100mM acetate, pH 5.2 and 500ul absolute ethanol. Incubate at -20°C for 20 minutes.

14) Pellet RNA by centrifugation at maximum speed in a microfuge for 10 minutes.

15) Wash the RNA once in 70% ethanol and vacuum dry.

16) Dissolve RNA in appropriate buffer i.e. DEPC-treated, sterile TE, pH 8 or 0.5% SDS if no enzymic manipulation of the RNA is needed. SDS is an inhibitor of ribonucleases. Total RNA was sufficient for most RACE, probe and cDNA preparations. If not, poly-A purification and cap-selection were used to increase the amount of full-length message in the samples.

Poly-A RNA

Poly-A RNA was purified from 10 µg aliquots using oligo-dT Sephadex (Oligodex) columns and standard techniques (Sambrook). Final yield was typically less than 100 ng per sample. This was later attempted with commercial Poly-A RNA (Clontech) with identical results.

The general protocol.

Before using oligo-dT bound column material, it is best to wash the binding medium to remove fine particle of latex. To wash the latex, transfer appropriate amount of beads/media (300 ul of the suspension of Oligotex per 1mg of total RNA) into a microfuge tube. Spin for 3 min at 12,000 rpm. Discard the supernatant. Gently suspend the latex in the same volume of 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.1% SDS. Spin again. Gently suspend the latex in the same buffer.

1. Add 1 mg of total RNA dissolved in RNase-free water to 300 ul of the Oligotex-dT30 suspension.

2. Incubate for 3 minutes at 65 C. Chill on ice.

3. Add 0.2 volume of 5M NaCl. Incubate for 10 minutes at 37 C.

4. Centrifuge for 3 minutes at 15,000 rpm. Discard the supernatant.

5. Suspend the pellet in 1 ml of washing buffer (10mM Tris-HCl pH7.5, 1mM EDTA, 0.5M NaCl, 0.1% SDS).

6. Centrifuge for 3 minutes at 15,000 rpm. Discard the supernatant.

7. Suspend the pellet in 300 ul of RNase-free water containing 0.1% SDS.

8. Incubate for 5 minutes at 65 C. Chill on ice.

9. Centrifuge for 3 minutes at 15,000 rpm. Transfer the supernatant into new microfuge tube.

10. Carry out phenol-chloroform extraction and ethanol precipitation by standard procedure. Rinse the pellet with 75% ethanol. Dissolve the poly(A)+ RNA in 10 ul of RNase-free water.

Cap-selection of RNA and 5' RACE

Here Calf Intestinal Phosphatase (CIP) is used to remove 5' phosphates from uncapped (either unspliced or degraded) RNA. A subsequent Tobacco Acid Phosphatase (TAP) reaction then cleaves the 7-Methyl-Guanidine capped (full-length) messages. This now leaves a 5' mono-phosphate. The RNA linker is then ligated to the RNA pool. The adaptor will only ligate to the full-length message which has the phosphate necessary for the reaction. Now an RT-PCR reaction is performed using thermo-labile high fidelity

reverse transcriptase (Superscript III, Invitrogen) and primers specific to the known 5' end of the message and the RNA linker.

The general protocol:

1. 500 ng poly(A) RNA is used in a CIP reaction with 2 µl buffer, 2 µl CIP (10 u) and water to 20 µl. 37° for one hour.

2. Add 15 µl 3M Ammonium Acetate pH 5.2, 115 µl dwater, 150 µl acid phenol:chloroform. Vortex, centrifuge 5 min. 13k g, extract top layer, add 150 µl chloroform, spin, 5 min 13k g, extract top layer, add 150 µl isopropanol, chill to precipitate, centrifuge 20 min. 13k g at room temperature, rinse with 500 µl chilled 70% ethanol, centrifuge 5 min. 13K g. Re-suspend in 8 µl 1x TAP buffer.

3. TAP reaction: 8 µl RNA-in TAP buffer, 2µ TAP. 37°, 1 hour.

4. Ligation reaction: 2µl CIP-TAP treated RNA, 1µl RNA Adaptor, 1µl buffer, 2 µl T4 RNA ligase, 4µl dWater. 37° 1 hour.

RNA Adaptor-

5'GGGUUCGGGCUUAGGCUCCAGUGCCUGUUCGGUGGUCGCGGCGCUGAUG GCGAUGAAUGAACACUGCGGCAAGCCGCUUAAUGACACUCGUUUGCUGGC UUUGAUGGGCGAGCUGGAAGGCCGUAUCUCCGGCAGCAUUCAUUCAUUUA CGACAAA-3'

5. Reverse Transcription Reaction: 2µ ligation reaction, 4µl dNTP mix (10µM each), 10 µM gene specific primer (or Random Decamers), 2µl 10x buffer, 1µ RNaseOut RNase inhibitor, 1µl Superscript III (or Thermoscript) in a 20 µ reaction.

We incubated as high as 55°C with a prior heat denaturation of the ligation product at 70°C. This was necessary to minimize secondary structure that inhibits the progress of the reaction.

6. Nested PCR: two pairs of primers were designed (actually multiple sets were required since the adaptor and inner gene primers are extended with a BamH I cut site) for cloning purposes, which makes primer design difficult to predict. It was later discovered that shorter primers without cuts sites were more effective. The PCR reaction were conducted with conditions suitable for the primers, but typically, a range of annealing temperatures and Mg ion concentrations were used. In all cases, 1µl of RT reaction was used.

NUP Primer 5'-AAGCAGTGGTAACAACGCAGAAGAGT-3'

A dilution of 1:100 of the first PCR reaction is used as a template for the second nested reaction.

TAP-minus RNA pool was also used in a ligation-RT-PCR series as a control to ensure that products of the nested PCR were from capped message only.

Gene-specific primers:

Kv3.2:

Outer: 5'-CGGCGTTATGCTGTGCTGATGAT-3'

Inner: 5'-TGCTGATGATACTGATGGGTCTGCGGTGGTTCTCC-3'

DNA Preparation

Standard DNA preparation techniques were employed along with some modifications to suite the needs of certain procedures. Drosophila present a particular difficulty in that the

proportion of protein to nucleotides is sufficiently great to necessitate additional steps in the process. In fact, several different methods were used depending upon the species.

Genomic

Welcome Bender-Jay Hirsch-Nigel Atkinson Method

Flies frozen in liquid nitrogen are ground in a ceramic mortar at -70 C and the resulting powder is added to a 2 ml Dounce homogenizer with 1 ml of homogenizing solution.

Fly homogenizing buffer:

0.1M NaCl

0.2M Sucrose

0.01 Na_2EDTA

0.03M Trisbase pH8.0

Transfer to autoclaved 15 ml corex tubes. 65°C 30 minutes

add 300 µl 8M KOAc, mix well, 0° C for 60 minutes

Spin 8K 8 minutes at 4° C (SS34 or equivalent)

Pour supernatant into 2 microfuge tubes (1.5 ml)

Spin 10 minutes microfuge at 4° C

Keep supernatant and divide into 3 microfuge tubes

Add 700 µl of room temperature 100% ethanol to each tube

Let stand 5 minutes at room temperature; then 5 minutes spin in 4° C microfuge

Wash pellet with cold 70% ethanol and re-suspend in 100 µl TE, pool all 3 tubes

Add 6 µl of 5M NaCl (final conc is 0.1 mM) and 15.6 µl of 200 mM spermine 4HCL

Spin 10 min, 13K g; remove supernatant

Re-suspend pellet in 180 µl of H20

Add 20 µl (1/10th volume) 3M NaOAc pH 5.5 and 500 µl ethanol

-20 □C for 30 minutes to o/n; spin 10 minutes 13K g

Wash pellet 2 times with cold ethanol

Atkinson's combined chemical lysis-double phenol-Qiagen prep

Homogenize flies as above, then suspend powder into 30 mls of rapidly stirring ice-cold NIB. Transfer to ice-cold 30 ml corex tube. Spin briefly at low speed to pellet large mass material.

Transfer to a new ice-cold corex tube and spin 7000 rpm 4°C 7.5 minutes.

Re-suspend pellet in 19 ml ice-cold NIB and transfer to 50 ml conical tube. Add 8 ml 10% Sarkosyl. Mix slowly by inverting. Let stand on ice 10 minutes.

This is the second departure point, normally this would lead to a CsCl prep, but we precipitated the chromatin and brought that material up in P1 of the Qiagen prep.

Nuclear Isolation Buffer:

37.5 mM Tris [pH 8.5], 0.05 mM spermine, 0.125 mM spermidine, 0.5 mM EDTA [pH 7.4], 20 mM KCl, 0.5% thiodiglycol, 0.05% Empigen BB, 0.1 mM PMSF, (2 µg of aprotinin per ml)- the pH is different from Bingham, Levis and Rubin.

Plasmid

With the advent of Qiagen kits (Qiagen, Venlo, Netherlands) alkaline lysis preps of large or medium scale plasmids have been relegated to the past. However, for extremely clean preparation of large quantities of DNA suitable for injections, we found that the alkaline-lysis prep below produced the best results:

Alkaline lysis 50 ml midi-prep

Spin down a 16-20 hour 50 ml culture, 5K g 30 minutes.

Add 3.0 ml Solution 1; vortex to suspend pellet; let stand on ice 10 minutes

Add 6.0 ml Solution 2; do NOT vortex- mix by inverting slowly; let stand on ice, 10 min.

Add 4.5 ml Solution 3; mix by inverting slowly; let stand on ice 10 minutes

Spin, 5K g 30 min.

Pour through Kimwipe into a fresh 50 ml conical.

Add 11 ml isopropanol; precipitate pellet (30' or overnight)

Re-suspend pellet in 200 µl TE; add 200 µl 5M LiCl; let stand on ice 10 minutes

Spin, 13K 10 minutes, draw off supernatant and transfer to a clean 1.5 ml tube

Precipitate with 400 µl isopropanol (30 min. or overnight), spin 13K g 10 min.

Re-suspend in 500 µl TE and treat 10 min. with 4 µl RNase (10 µg/µl) 37° C, 15 min.

Phenol-chloroform extract, precipitate o/n with NH_3oAc and EtOH.

Yields 100-400 µg DNA

Solution 1: 1% glucose, 25 mM Tris-HCl, pH 7.5, 10 mM EDTA

Solution 2: 0.2 N NaOH, 1% SDS

Solution 3: 3 M KoAc

Cosmid/PAC/BAC prep

This is adapted from Roe (Univ. of Wash), which is derived from Chen, Pan, and Ying at the Berkeley Genome Sequencing Centre. This employs a double acetate precipitation. These are low copy number vectors, and so require staged growth:

Single colonies are streaked and grown overnight. This smear is then used to incubate a 50 ml culture that is grown for 8-10 hours. This culture is then transferred to a 250 ml

flask with an equal volume of the same medium. This is grown over night for 12-16 hours and 50 mls is used to inoculate a 1 L culture in a 3 L flask. This is divided into 500 ml bottles and spun down. If pellets are frozen overnight at -70° C, there is a higher yield.

In Situ Hybridization

This protocol describes a procedure for the localization of mRNA with whole mount Drosophila embryos. It was also adapted for use with larval and adult brains. It is a modification of the Vectastain ABC labeling kit (VectorLabs, Burlingame, Ca). A digoxigenin labeled probe is prepared and then hybridized to whole mount specimens. A secondary antibody against digoxigenin conjugated to alkaline phosphatase is hybridized and a chemical reaction is used to show color for localization.

Preparation of the probe

Total RNA from 100 Canton-S flies was prepared, then an RT-PCR reaction performed using Superscript 2 and PlatinumTaq, with random hexamers and the primer pairs:

Shaw Upper 1

Upper: 5' – CTGATCAACATGGACTCGGAA– 3'

Shaw Ex1

Lower: 5' – AAACCGTCCGAAGAGGAATT – 3'

Shaw2 U6646

Upper: 5' – TAACTGCTGGCAGCGTATAAAA– 3'

Shaw2 L7408

Lower: 5' – CATAAAGTCGCCGGTCAATATT– 3'

These produced ~500 bp probes that were then twice gel-purified. Starting with 400 ng in 15 µl the probes were labeled with random primers and alkali-labile dUTP-Digoxigenin conjugate for 19 hours by adding 2 µl of random hexanucleotide primers; 2 µl of dNTP labeling mix (Roche [details]) and 1 µl Klenow enzyme. They were ethanol

precipitated; brought up in 40 µl water and quantified. A final concentration of ~24 ng/µl (~1000 ng total) was generated for each probe. Of this, 45 µl of probe diluted in 155 µl of hybridization buffer (50% deionized formamide, 5X SSC, 100 µg/ml ssDNA, 200 µg.ml tRNA, and 0.1% Tween-20) was used in the in situ.

Preparation of whole mount material

Late stage (18-21 hour) embryos were collected from egg plates, rinsed with water and dechorionated for 2.5 minutes in 50% bleach and rinsed again with PBS, 0.3% Triton-X. These were transferred to a 1.5 ml microfuge tube containing 50% heptane and 50% fixative (3.7% formaldehyde, 50 mM EGTA in PBS (Phosphate Buffer Saline, 10x: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO4, 2 mM KH_2PO_4) and fixed for 20 minutes on a rocking platform. The lower formaldehyde phase is removed and 1 ml of methanol is added to clear the embryos and remove the vitelline membrane. Shake vigorously for 15 seconds (we vortexed embryos for 30 seconds also) and let stand for 1 minute. Then remove the upper heptane layer and add more methanol. Repeat the methanol washes four times. Then wash four times in ethanol. Embryos are stored at -70° C to reduce background. They can be collected over time to accumulate enough material for the assay. They are then rehydrated in stages: 1 minute each step: 25% PBS-T (1X PBS, 0.1% Tween-20)/MeOH; 50% mix; 75-25% Mix.

We fixed the embryos again at this point; washed in PBS-T and incubated at 37° C for ~3 minutes with non-predigested proteinase-K (this is after running samples at different time points to evaluate the efficacy of the Proteinase-K). Stop the Proteinase-K activity with 2mg/ml glycine in PBS-T. Too long and the embryos disintegrate. Post-fix in 5% formaldehyde and 0.2% gluteraldehyde for 20 minutes. Then five 2 minute washes in

PBS-T. Then remove 50% of PBS-T and add 50% hybe solution (50% deionized formamide, 5XSSC, 100 µg/µl ssDNA, 200 µg/µl tRNA, 0.1% Tween-20) to a 1:1 mixture. Then remove this and add 100% hybridization solution. Pre-hybridize 48° C for 2 hours.

Hybridization

Add 1 µg/ml heat denatured digoxygenin labeled probe (boil 5 minutes and then quick chill on ice). Hybridize at 48° C for 24-36 hours. Wash embryos in staged reduction of Hybridization solution (25% PBS-T: 75% Hyb. solution; 50% each; then 75% PSB-T: 25% Hyb. solution). Perform two washes of 20 minutes each with PBS-T. Alkaline Phosphatase-conjugated IgG Fraction Monoclonal Mouse Anti-Digoxin (Code# 200-52-156, JacksonResearch, West Grove, PA) is added at 1:2000 dilution of stock (600 µg/ml). Incubate for one hour and then wash four times in 20 minutes each in PBS-T.

Chemical Reaction

The native alkaline phosphatase activity must be blocked. Wash twice in 100 mM NaCl, 50 mM $MgCl_2$, 1 mM Levamisol (Sigma, L9756-5G), 0.1% Tween-20, for two minutes each. Add to the second 1 ml wash: 4.5 µl NBT and 3 µl X-Phosphate (BCIP, Roche, 10742020). Rock 5 minutes to several hours and watch as color develops. Stop reaction by washing at least six times in PBS-T. Wash once in 40% glycerol, then twice in 80% glycerol.

Bioinformatic Analysis

With the release of whole genome sequences, there has been an explosion of the number of tools available for analysis of these sequences— either in toto or with specific patterns in mind. These programs fall into three general categories: annotation tools such as national databases that allow for rudimentary searches and submission of descriptive information; pattern analysis tools that search the sequences for matches against specific external sequences of known importance; and tools that look for internal patterning of sequences (information content) that may have been preserved through time due to some functional constraint.

Software

Databases

Eventually, all sequences were housed on local machines for analysis, however these were regularly reconciled with updated databases available on the Internet.

Flybase

This is a clearing house of information regarding Drosophila that is primarily housed at the University of California at Berkeley, but is mirrored at Harvard University, Indiana University, and in Europe at Cambridge University.

http://flybase.net/

The specific entries for the genes in this study are:

Kv3: http://www.flybase.org/.bin/fbidq.html?FBgn0003386

Kv3.2: http://www.flybase.org/.bin/fbidq.html?FBgn0032113

Transfac

Transfac is a database and search engine that utilizes curated transcription factor binding site data sets in discrete and matrix form. Raw sequence is matched against these databases with either the basic blast (Patch: the PAttern maTCH section) or a log-odds scoring matrix algorithm (Match: the MAtrix maTCH section). Transcription factor binding sequences are assigned a 'core' region of 5 base pairs that can be matched at a different stringency than the entirety.

This is a commercial project that is also available publicly as a subset of the total number of entries available to subscribing customers. With a subscription, the database and search engine are run locally. This allowed for the entry of additional matrices into the database. The Biobase (7.4 version) contained 38 insect entries. This was expanded to 62 locally using published data.

The public and commercial site is: http://www.gene-regulation.com/

Several other genome search algorithms now link sequence comparisons to the public version of this database (for example rVista).

On-line data bases.

There is no shortage of available databases on the internet. At this point, the list is expanding to the extent that any summation here would be futile. The reader is encouraged to use google.

Sequence Analysis

ClustalW

The source code for clustalW was obtained and modified and recompiled to accommodate the expanded symbols list used in the production of pseudosequences from transfac runs. The matrices were re-written with upper and lower case letters assigned to transcription factor binding site search results. A new identity matrix was created and linked to the runtime library.

Otherwise, the unmolested version of clustal 1.83 was used.

Macvector

Sequence annotation tool. Simple searches and graphical output of annotated sequences. Available commercially at an exhorbitant price.

Genomic Alignements

A resource for aligning whole genomes is available online at:

http://pipeline.lbl.gov/cgi-bin/gateway2

This in turn, is tied to a similar clearing house of genomic annotations at the University of California at Santa Cruz.

References

Adams, MD *et al*. The genome sequence of Drosophila melanogaster. Science. 2000;287(5461):2185-2195.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular biology of the cell, fourth edition. New York: Garland Science, 2002.

Attali B, Lesage F, Ziliani P, Guillemare E, Honore E, Waldmann R, Hugnot JP, Mattei MG, Lazdunski M, Barhanin J. Multiple mRNA isoforms encoding the mouse cardiac Kv1-5 delayed rectifier K+ channel. J Biol Chem. 1993 Nov 15;268(32):24283-9.

Baro, DJ *et al*. Shab gene expression in identified neurons of the pyloric network in the lobster stomatogastric ganglion. Receptors Channels. 1994;2(3):193-205.

Biessmann H *et al*. Euchromatic and heterochromatic domains at Drosophila telomeres. Biochem Cell Biol. 2005;83(4):477-485.

Butler A, Wei A, Baker K, Salkoff L. A family of putative potassium channel genes in Drosophila. Science. 1989;243:943-947.

Dermitzakis ET, Bergman CM, Clark AG. Tracing the evolutionary history of Drosophila regulatory regions with models that identify transcription factor binding sites. Mol Biol Evol. 2003a;20(5):703-714.

Dermitzakis, ET *et al*. Evolutionary discrimination of mammalian conserved non-genic sequences (CNGS). Science. 2003b;302(5647):1033-1035.

Draper, M. Transcriptional regulation of the shaker homolog Kv3. PhD dissertation, University of Texas at Austin, 2005.

Gregory, SG *et al*. A physical map of the mouse genome. Nature. 2002;418(6899):743-750.

Hille B. Ionic channels of excitable membranes, 2d ed. Sinauer, Sunderland, Mass, 1992.

Hodge JJ, Choi JC, O'Kane CJ, Griffith LC. Shaw potassium channel genes in Drosophila. J Neurobiol. 2005 Jun;63(3):235-54.

Jiang Y MacKinnon R *et al*. Structure of the RCK domain from the E. coli K+ channel and demonstration of its presence in the human BK channel. Neuron. 2001 Mar;29:593-601.

Kamb A, Iverson L, Tanouye MA. Molecular characterization of Shaker, a Drosophila gene that encodes a potassium channel. Cell. 1987;50:405-413.

King MW. Control of gene expression. 22 Mar 2006. Indiana University School of Medicine. 1 Sept 2006 <http://www.indstate.edu/thcme/mwking/gene-regulation.html>.

Kraft P, Horvath S. The genetics of gene expression and gene mapping. Trends Biotechnol. 2003;21(9):377-378.

Levine M, Tijan R. Transcription regulation and animal diversity. Nature. 2003;424(6945):147-151.

Mottes JR, Iverson LE. Tissue-specific alternative splicing of hybrid/lacZ genes correlates with kinetic differences in Shaker K+ currents in vivo. Neuron. 1995;14:613-623.

Ohler U, Liao GC, Niemann H, Rubin GM. Computational analysis of core promoters in the Drosophila genome. Genome Biol. 2002;3(12):RESEARCH0087.

Rhettig J, Wunder F, Stocker M. *et al*. Characterization of a Shaw related potassium channel family in rat brain. EMBO J. 1992;11:2473-2486.

Rubin GM *et al*. Comparative genomics of the eukaryotes. Science. 2000;287(5461):2204-2215.

Shahidullah M, Harris T, Germann MW, Covarrubias M. Molecular features of an alcohol binding site in a neuronal potassium channel. Biochemistry. 2003 Sep 30;42(38):11243-52.

Smith-Maxwell CJ, Ledwell JL, Aldrich RW. Role of the S4 in cooperativity of voltage-dependent potassium channel activation. J Gen Physiol. 1998 Mar;111(3):399-420.

Tu L, Santarelli V, Sheng Z, Skach W, Pain D, Deutsch C. Voltage-gated K+ channels contain multiple intersubunit association sites. J Biol Chem. 1996 Aug 2;271(31):18904-11.

Ventor JC *et al*. The sequence of the human genome. Science. 2001;291(5507):1304-1351.

Wei A, Covarrubias M, Butler A, Baker K, Pak M, Salkoff L. K+ current diversity is produced by an extended gene family conserved in Drosophila and mouse. Science. 1990;248:599-603.

Wiehe T, Guigo R, Miller W. Genome sequence comparisons: Hurdles in the fast lane to functional genomics. Brief Bioinform. 2000;1(4):381-388.

Xu, J. *et al*. Assembly of voltage-gated potassium channels: Conserved hydrophilic motifs determine subfamily-specific interactions between α-subunits. J Biol Chem. 1995;270:24761-24768.

Yu JY, Upadhyaya AB, Atkinson NS. Tissue-specific alternative splicing of BK channel transcripts in Drosophila. Genes Brain Behav. 2006 Jun;5(4):329-39.